Interoperability Session

BRC Kick-off Meeting
Oct 13-14 2004

Bruno Sobral (sobral@vbi.vt.edu)
Virginia Bioinformatics Institute
http://www.vbi.vt.edu/

VIRGINIA BIOINFORMATICS INSTITUTE AT VIRGINIA TECH

Agenda

- **Presentation by Valentina Di Francesco (NIAID)**
- **Presentation by me**
- **Round Table Discussion on Topics Sent Out**
- **Break**
- **Round Table Discussion: Establishing of BRC IOWG**

Agenda

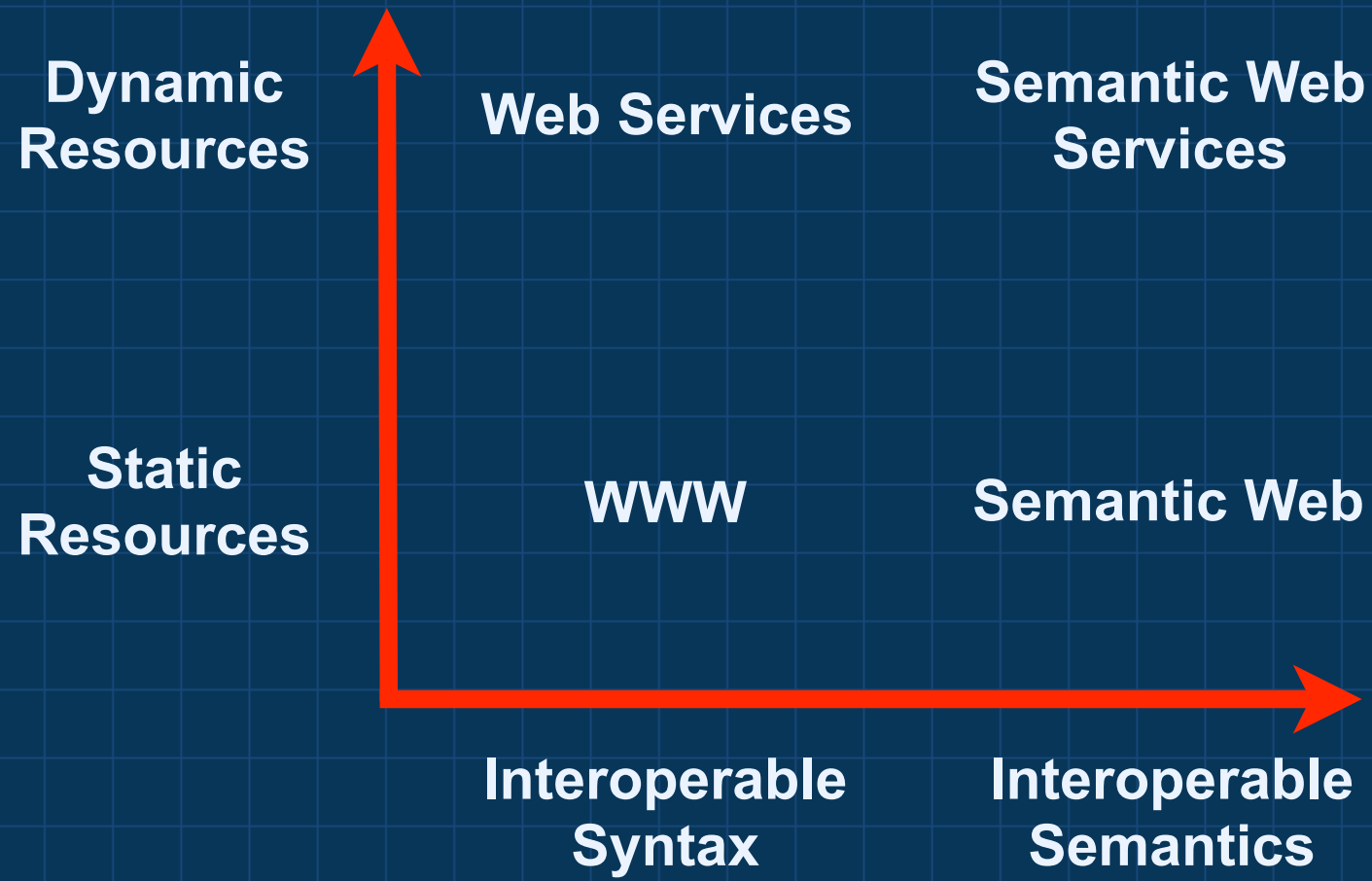
- Presentation by Valentina Di Francesco (NIAID)
- **Presentation by me**
- Round Table Discussion on Topics Sent
- Break
- Round Table Discussion: Establishment of BRC's IOWG

What Is Meant By Interoperability?

- **"...suitable for and capable of being implemented in a neutral manner on multiple operating systems and in multiple programming languages." (WS-I)**
- **Deals with syntax, *not* semantics (see Host-Pathogen Interaction Session, under Ontology)**
  - **Syntax *required* to achieve integration (needs semantics as well)**
- **Focus: establishing machine-readable interfaces for key BRC components**

VIRGINIA BIOINFORMATICS INSTITUTE AT VIRGINIA TECH VBI

Semantic Web Services

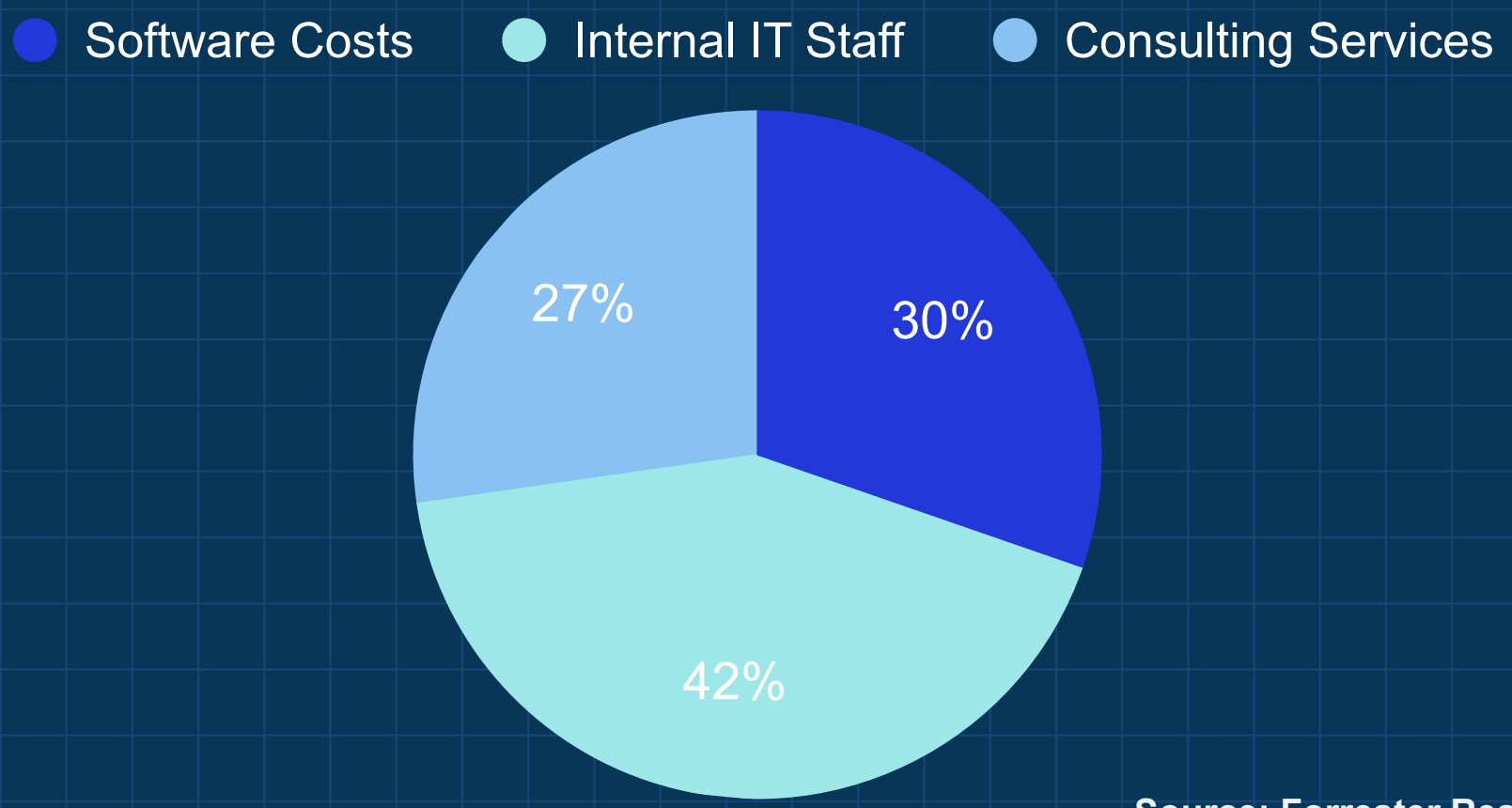


Source: Daconata et al., *The Semantic Web*

Why Does Interoperability Matter to BRCs?

- Imagine needing to do [comparative genomics] across bacterial (or viral) species whose data are contained in multiple BRCs....
- Imagine trying to integrate great (software) components from other BRCs....
- Imagine trying to integrate BRC systems if they are built as "proprietary" individual systems....
- Imagine trying to analyze host responses to diverse pathogens in which data are distributed across BRCs (and RCEs)....
- Etc....

Lack of Interoperability Is Costly: Average Cost/Business in 2001 = $3.5 Million



Source: Forrester Research

Cyberinfrastructure Characteristics

Cyberinfrastructure Characteristics

- Community-Focused
  - virtual organizations
  - distributed,
  - collaborative
- Scale and Scope
  - Multidisciplinary
  - International
  - Supporting data- and compute-intensive applications
  - High-end to desktop
  - Heterogeneous
- Common Technology & Policy Platform(s)
  - Interoperability
  - Supports characteristics above

VIRGINIA BIOINFORMATICS INSTITUTE AT VIRGINIA TECH VBI

Debbie Crawford, NSF

(Some of) The Challenges

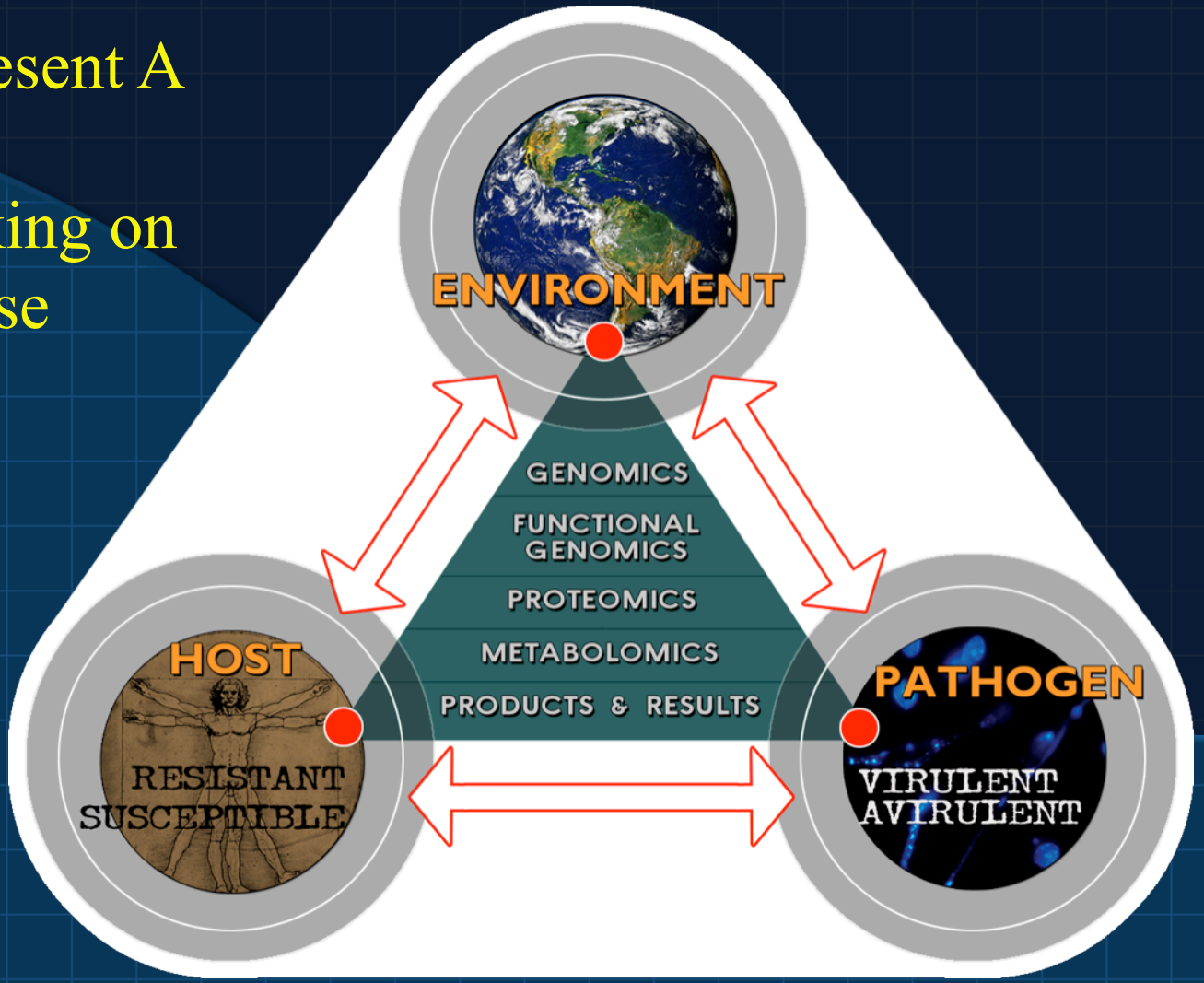
Challenging Context

- Institutional & Infrastructural Ecology
  - Technological change more rapid than institutional change
- Broadening Participation
- Community-Building
- Seamless Integration of New and Old
  - Balancing upgrades of existing and creation of new resources
  - Legacy data/models
- Providing sustainable support

VIRGINIA BIOINFORMATICS INSTITUTE AT VIRGINIA TECH VBI

PathoSystems Biology

The BRCs Represent A Component of a Federation Working on Infectious Disease Research



Role of IT in PathoSystems Biology

VIRGINIA BIOINFORMATICS INSTITUTE AT VIRGINIA TECH

BIOTERRORISM

Information Technology Strategy Could Strengthen Federal Agencies' Abilities to Respond to Public Health Emergencies

GAO Report, May 2003, 03-139

What GAO Found

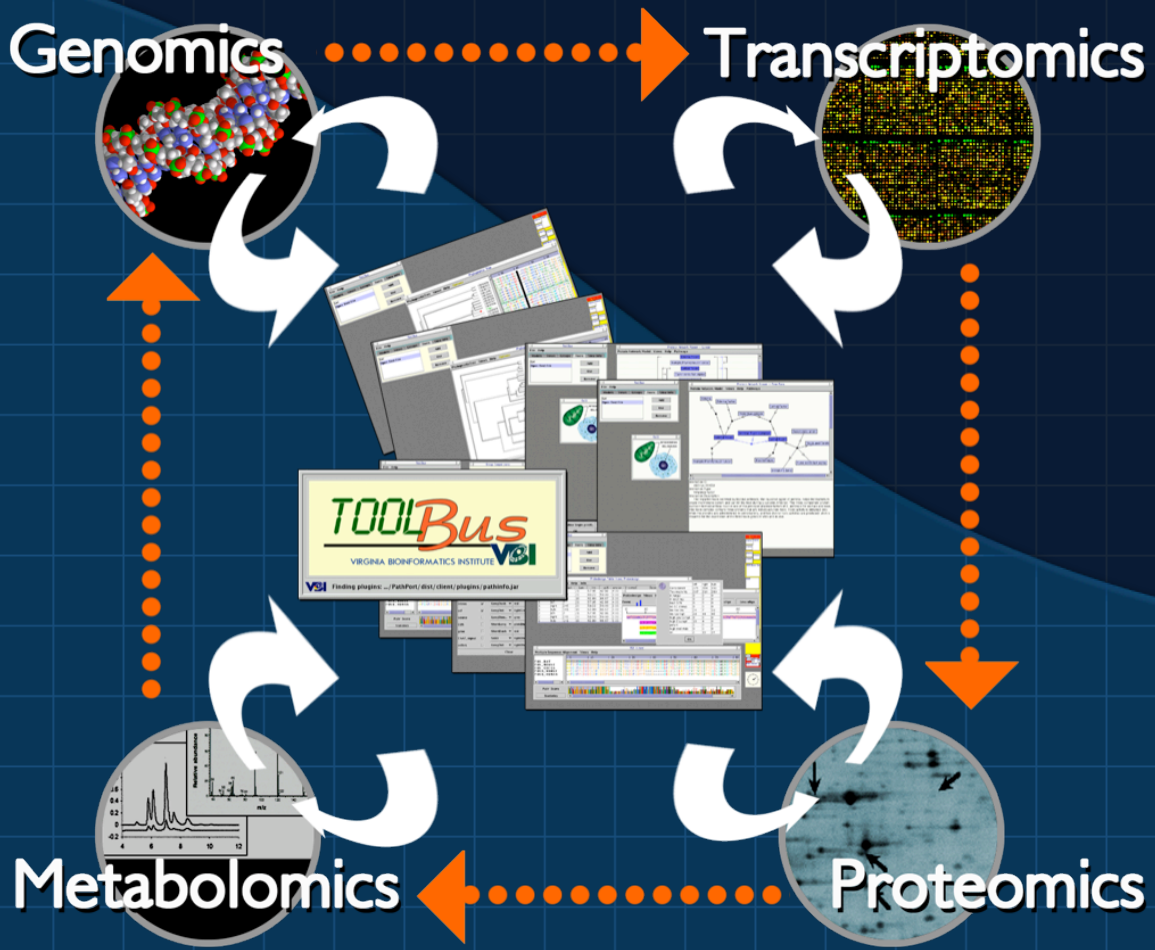
The six key federal agencies involved in bioterrorism preparedness and response identified about 70 planned and operational information systems in several IT categories associated with supporting a public health emergency. These encompass detection (systems that collect and identify potential biological agents from environmental samples), surveillance (systems that facilita
related
deliver
suppor
other c

Summary of the Systems Inventory by Agency

IT Categories	HHS	Defense	Energy	Agriculture	EPA	VA	Total
Detection	0	4	6	0	0	0	**10**
Surveillance	18	7	2	6	0	1	**34**
Communications	5	2	0	3	0	0	**10**
Supporting Tech	5	1	6	1	5	0	**18**
Total	**28**	**14**	**14**	**10**	**5**	**1**	**72**

Source: GAO.

Systems Interoperation: An Anecdote from the Pathogen Portal Project



VIRGINIA BIOINFORMATICS INSTITUTE AT VIRGINIA TECH VBI

Build on Open Standards

- Common vocabulary: Gene Ontology (GO)
- Transport format: XML
- Data definition language: XSD
- Wire protocol: SOAP
- Service definition language: WSDL
- Service registry: UDDI

After L. Stein, CSHL

PathPort's Use of Community XMLs

- Utilizes established, open community standards
  - ***DAS-ML, BSML, MSA-ML (DNA) - Year 1***
  - ***MAGE-ML (mRNA profiling) - Year 2***
  - *PEDRo (protein profiling) - Year 3*
  - *SBML (molecular models) - Year 3*
  - CellML (cellular levels including metabolism and signal transduction) - Year 4
  - AnatML (organ levels) - Year 4
  - FieldML (spatially and temporally varying field information using finite elements) - Year 5

VIRGINIA BIOINFORMATICS INSTITUTE AT VIRGINIA TECH VBI

ToolBus

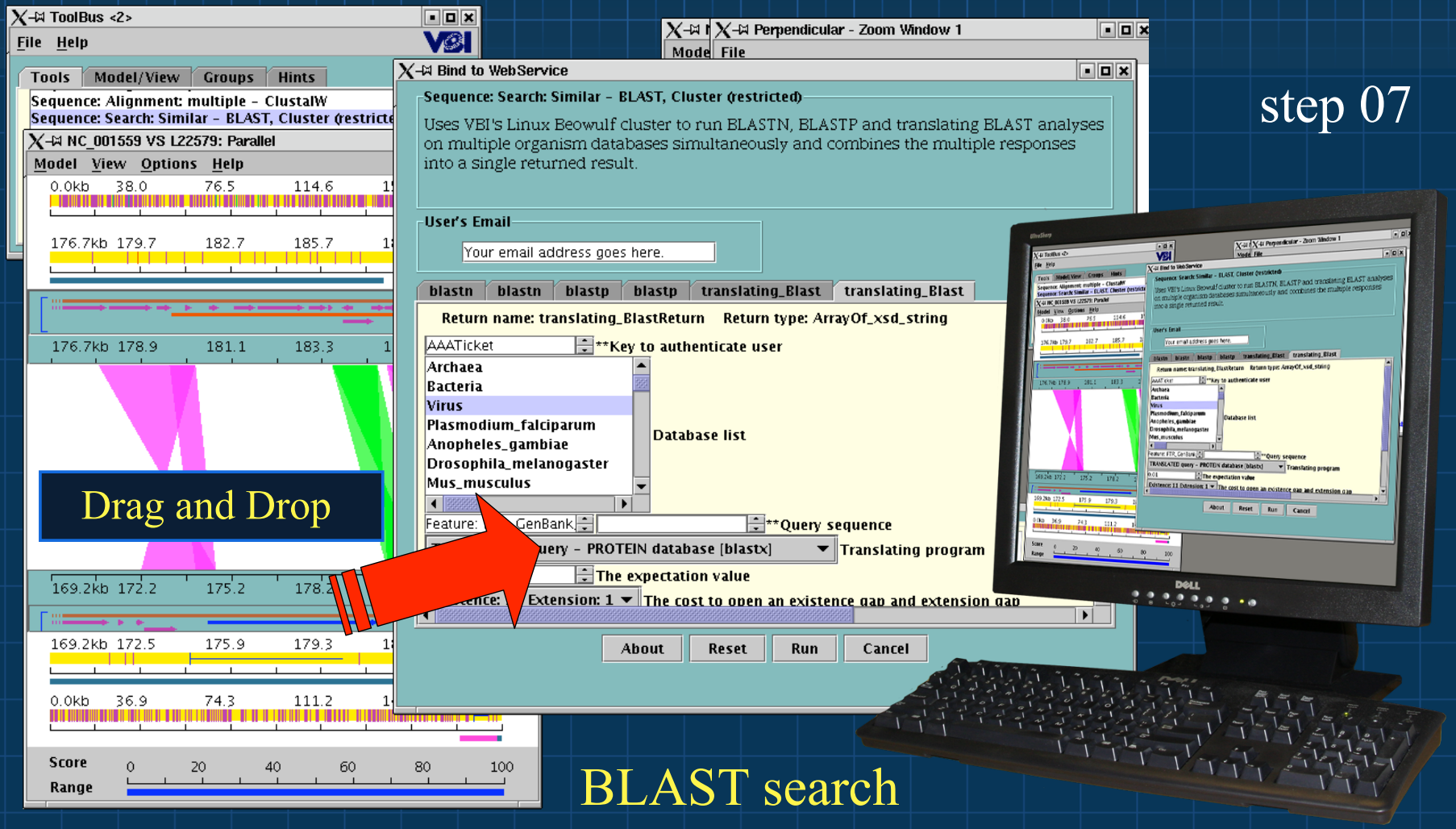
A client-side interconnect with the following goals:

- Platform independent
- Easily extensible
- Allow user-defined associations
- Easy to use

PathPort Comparative Genomics Scenario

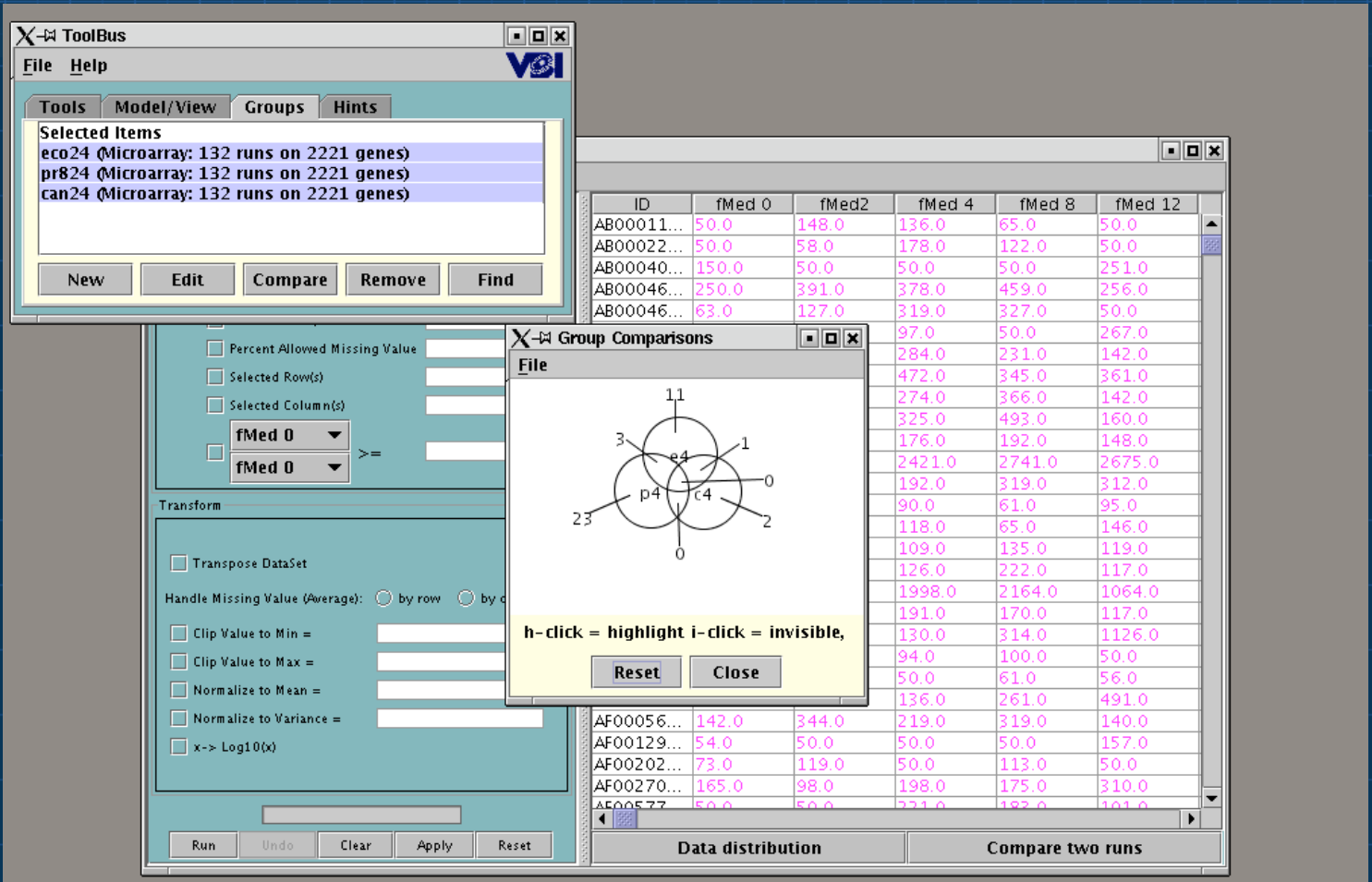
VIRGINIA BIOINFORMATICS INSTITUTE AT VIRGINIA TECH

step 07



BLAST search

PathPort Group Comparison Scenario Seeking Diagnostic Markers



Comparison of three groups

SCRUM: Focus on Product, not Process

PathPort SCRUM Web Site

Admin Login
About Scrum
Suggest Feature
Search
BackLog Stats
BackLog Items
Completed
Released

Sprints
- 30 Jan 2004
- 09 Jan 2004
- 31 Dec 2003
- 21 Nov 2003
- 03 Nov 2003
- 30 Oct 2003
- 08 Oct 2003
- 07 Oct 2003
- 03 Oct 2003
- 01 Oct 2003
- 30 Sep 2003
- 03 Sep 2003
- 02 Sep 2003
- 28 Aug 2003
- 27 Aug 2003

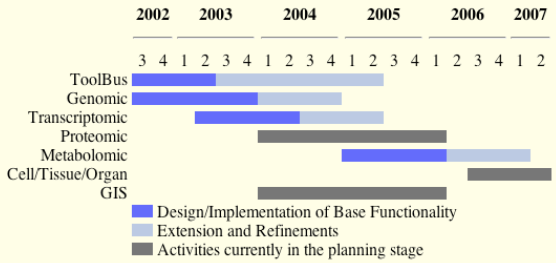
Help us make *ToolBus/PathPort* a better system!

Note: Bug reports should be reported here instead.

As a *PathPort* user you know better than anyone what changes, additions, modifications, and general improvements will make your use of the system more productive. "Suggest Feature" allows you to tell us what *ToolBus/PathPort* needs to make it better. PLEASE be sure to include your email address in case we have questions regarding the details of your suggestion.

In addition, if you'd like to see the progress we've made or would like to see the suggestions that other users have made, please feel free to look around.

Five Year PathPort Work Plan



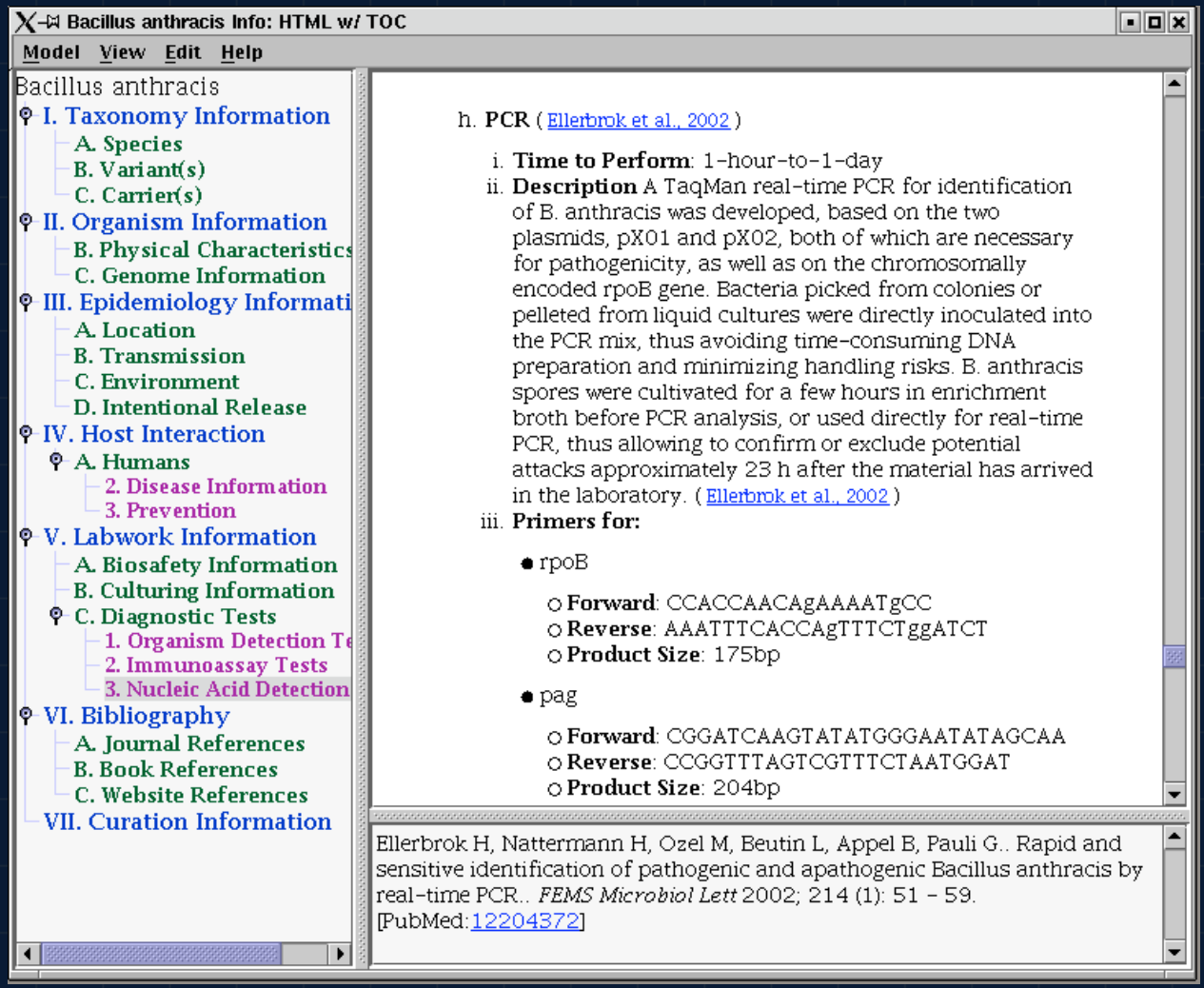
NOTE: The above "timebars" are clickable, but are NOT broken down by quarter.

for more on SCRUM: Schwaber and Beedle - *Agile Software Development with Scrum*

SCRUM Provides Interface for Community Requests

Literature Curation: Pathogen Background Information

PathInfo Viewer



Pathogen Information Markup Language (PIML)

THE PATHOGEN PORTAL
https://www.vbi.vt.edu/pathport

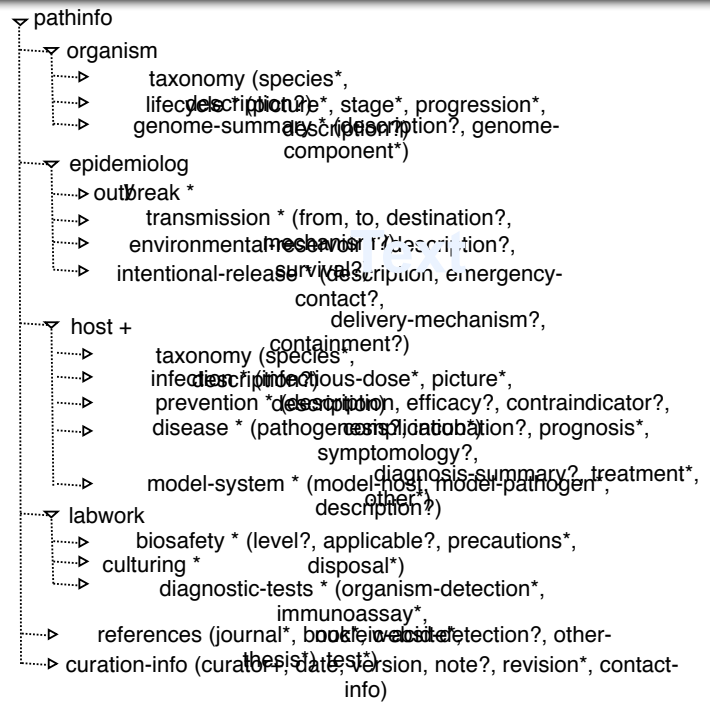
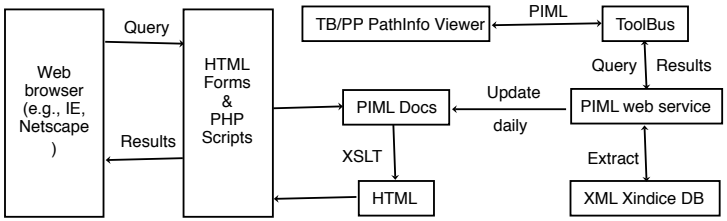


Figure 1. The PIML structure overview. The lower levels of elements and all element attributes are not shown but can be viewed from the PathInfo project website. A particular element can appear one or more times, zero or more times, or zero or one times, signaled by the plus sign (+), asterisk (*), and question mark (?), respectively.

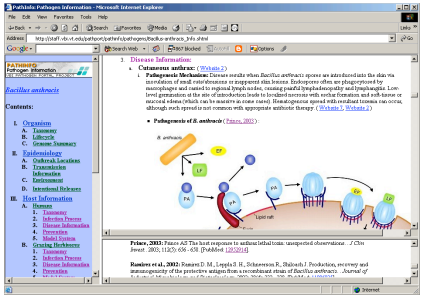
He *et al.*, Bioinformatics, in press

PathInfo Query and Display System

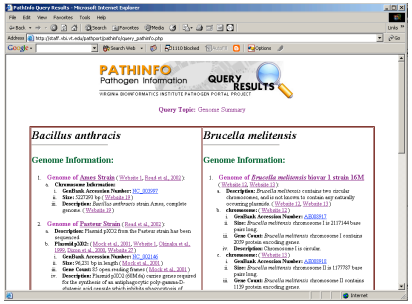
THE PATHOGEN PORTAL
https://www.vbi.vt.edu/pathport



(A) Architecture



(B) Complete Pathogen Information



(C) Comparison of Pathogens

He *et al.*, Bioinformatics, in press

Agenda

- Presentation by Valentina Di Francesco (NIAID)
- Presentation by me
- **Round Table Discussion on Topics Sent**
- Break
- Round Table Discussion: Establishment of BRC's IOWG

Agenda

- Presentation by Valentina Di Francesco (NIAID)
- Presentation by me
- Round Table Discussion on Topics Sent
- Break
- Round Table Discussion: Establishment of BRC's IOWG

Agenda

- Presentation by Valentina Di Francesco (NIAID)
- Presentation by me
- Round Table Discussion on Topics Sent
- Break
- Round Table Discussion: Establishment of BRC's IOWG